

UNCLASSIFIED

AD 422821

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

422821

nat

⑤ 739 200

①

422821

AD 10. FILE COPY

⑤ 739200

⑥ SUB-OPTIMIZATION IN OPERATIONS PROBLEMS,

⑩ by Charles Hitch,

P-326 ✓

18 November 1952

The RAND Corporation

1700 MAIN ST. • SANTA MONICA • CALIFORNIA

25
21

SUB-OPTIMIZATION IN OPERATIONS PROBLEMS^{1/}

Charles Hitch

The validity and therefore the usefulness of operations research depend upon the skill with which projects are designed and particularly upon the shrewdness with which criteria ("payoffs," "objectives functions") are selected. The criterion problem has been relatively neglected in operations research literature, and has apparently usually been "solved" in practice by assuming the first plausible payoff function which springs to mind; or if several spring to mind, by trying all and compromising (or letting a commander compromise) among the results of alternative computations. The problem is much too important for such casual treatment. Calculating quantitative solutions using the wrong criteria is equivalent to answering the wrong questions. Unless operations research develops methods of evaluating criteria and choosing good ones, its quantitative methods may prove worse than useless to its clients in its new applications in government and industry.

Levels of Optimizing

The optimal (or less ambitiously, good) solutions sought by operations research are almost always "sub-optimizations" in the sense that the explicit criteria used are appropriate to a low (or at least not the highest) level with which the researcher and his client are really concerned.^{2/} "Level" can be given a rigorous definition in formulating any particular operations problem, although the definition may sometimes contain an element of arbitrariness, depending on how the problem is formulated. Its meaning is clear enough in a general and intuitive way: the plant is a lower level than the multi-plant company of which it is a part; the company is itself a lower level than the national economy^{3/}; the Battle of the Atlantic in World War II was a lower level than the war as a whole.

^{1/} I am indebted to several of my colleagues at RAND, particularly Roland McKean, for valuable suggestions and criticisms; and to A. W. Marshall for the Mathematical Note which is appended.

^{2/} By a natural extension the term sub-optimization can be applied to any use of a partial or incomplete objectives function, as defined below. I use the term in its narrower sense in this paper.

^{3/} Whether the operations researcher is concerned with the higher level in this case depends upon the identity of his client.

Operations researchers have to sub-optimize (use low level criteria) because it is so frequently impossible, either in principle or more frequently in practice, to calculate the consequences of any given action in terms of the appropriate high level criteria. Let me illustrate with an example of universal familiarity--the problems of a family in budgeting its income for the purchase of various goods and services. Suppose that the head of the family has at his disposal an operations researcher (or, if you prefer, a whole operations research organization). What can operations research--quantitative common sense and scientific method--do to help the family get the most satisfaction in spending its income?

One possibility--which might occur to a mathematician or to a very theoretical economist--would be to attempt a full and complete optimization in a single calculation. Formally, this presents no difficulties. For the last 100 years economists have been assuming that each consumer has a "utility" (or preference) function which shows the (rank) value to him of all possible combinations of goods and services which he can buy on the market. The implications of this assumption have been worked out with great elegance. "All" that our operations researcher would have to do would be to write down the family's utility function, cost the goods and services available in the market, and compute the solution, maximizing the utility or "objectives" function subject to the income constraint. The solution, of course, would tell the family precisely how to spend each dollar of its income.

Any experienced operations researcher knows intuitively that this approach would prove to be sterile and hopeless. It would break down at the first step. We could not write down the family's general utility function because the family could not tell us what it was, and we could not conceivably derive it from any other source. It is as elusive as its counterpart in military operations research--which sometimes goes by the name of "military worth."

Rejecting this full optimizing calculation, what of a less ambitious nature can our operations researcher do? There are two things (although we may want to call only the first "operations research").

1. He can break the family's problem down and make quantitative sub-optimization analyses of some income allocations at lower levels--employing appropriate lower level criteria.
2. He can also provide background studies--not necessarily quantitative--which would enable the household to evaluate and use the sub-optimization results more "intelligently" or which would help it make good decisions in those sectors in which sub-optimization calculations aren't now possible for one reason or another.

In family budgeting I think it is clear that there are very satisfactory sub-optimizations possible for some allocations of income, less satisfactory ones for others, and none at all for others. The possible sub-optimizations fall into three classes:

- (a) Those in which there is what I shall call a "one-dimensional" objectives function. There is only one important objective or, if there are several, all can be reduced, in practice, to a single measure. If our family in sub-allocating its travel budget simply wants to get to New York by the cheapest means of transportation, and cares about nothing else, the objectives function is one-dimensional and the operations researcher can make a definite recommendation.
- (b) Those in which the objectives function is multi-dimensional, but all important objectives are subject to quantitative analysis. Thus, if our family values both cheapness and safety in getting to New York, but can't put a money value on safety, operations

research can evaluate both, but the family must "take all the facts" and exercise judgment in reaching a decision.

(c) Those which are partial in that objectives which are known to be important are left out because they are not subject to quantitative analysis. Thus, if our family thinks it attaches importance to the thrill of a first airplane trip, an experienced operations researcher might help it put this "thrill" factor in proper perspective, but he would not need computing equipment to do so.

There will be other income allocation problems confronting the household where no quantitative calculations are possible at all, and they will not necessarily be the less important ones. Suppose, for example, that the real problem is not mode of transportation to New York but whether to spend a vacation in New York, Boston, or Washington. Our operations researcher makes a preliminary reconnaissance of this problem and discovers that the family has two primary objectives--historical monuments and good food. It had chosen New York in the first instance simply because it was ignorant of the attractions of Boston and Washington in both these respects. Culinary and topographical research could be of very great assistance to the family in making the best choice among the three, but like the evaluation of the thrill of air travel, it would be essentially non-quantitative in character.

Examples of Criterion Problems

The criterion problem is simplest in industrial operations research. The measuring rod of money, which is appropriate for both earnings and costs, usually permits the use of a one-dimensional objectives function--the profits of the operation or (at a higher level) of the firm. There are, however, complications associated with uncertainty and time; as well as temptations to maximize or minimize physical quantities whose measurement is easy but whose relation to profits is obscure. Straight, hard thinking about criteria is therefore necessary even in industrial applications.

In military operations research the need^{1/} is usually much greater, for the problem is typically a sub-optimization of the complex third type. It may have multi-dimensional objectives functions as well as some important non-quantifiable factors, so that a great deal of judgment must be exercised in interpreting the results--either by the operations researcher in making his recommendations or by the commander in deciding whether and how to implement them.

Occasionally an obviously appropriate one-dimensional objectives function permits a neat, simple, and completely persuasive solution to be presented--even in military applications. But criteria which appear plausible or even obvious at first glance are quite likely to turn out to be traps for the unwary.

Let me take an example from Morse and Kimball^{1/} with which most operations researchers are familiar--the convoy problem discussed in chapter 3. I feel that I can use this example with impunity because, while it perfectly illustrates the dangers of sub-optimizing with partial objectives functions, I am confident that the action based on the study happened to be right--thanks, doubtless, to the good sense and judgment of both the operations researchers and the commanders involved.

The data revealed that, over a wide range, the number of merchant vessels sunk in a U-boat attack on a convoy was proportional to the number of U-boats in the attacking pack and inversely proportional to the number of destroyer escorts, but independent of the size of the convoy. They also revealed that the number of U-boats sunk per attack was directly proportional both to the number of attacking U-boats and the number of defending escorts. The objectives function was taken (plausibly) to be the "exchange rate" or ratio of enemy losses (measured in U-boats) to our losses (measured in merchant ships).

^{1/} Philip M. Morse and George E. Kimball, Methods of Operations Research, The Technology Press of Massachusetts Institute of Technology and John Wiley and Sons, Inc., New York; Chapman and Hall, Ltd., London; 1951

I quote the conclusion: "The important facts to be deduced from this set of equations seem to be: (1) the number of ships lost per attack is independent of the size of the convoy, and (2) the exchange rate seems to be proportional to the square of the number of escort vessels per convoy. This squared effect comes about due to the fact that the number of merchant vessels lost is reduced^{1/}, and at the same time the number of U-boats lost per attack is increased^{1/}, when the escorts are increased, the effect coming in twice in the exchange rate. The effect of pack size cancels out in the exchange rate. From any point of view, therefore, the case for large convoys is a persuasive one.^{2/}

"When the figures quoted here were presented to the appropriate authorities, action was taken to increase the average size of convoys, thereby also increasing the average number of escort vessels per convoy. As often occurs in cases of this sort, the eventual gain was much greater^{1/} than that predicted by the above reasoning, because by increasing convoy and escort size the exchange rate (U/B sunk)/(M/V sunk) was increased to a point where it became unprofitable for the Germans to attack North Atlantic convoys, and the U-boats went elsewhere. This defeat in the North Atlantic contributed to the turning point in the 'Battle of the Atlantic.'"

This happy outcome depended on the intuition and good sense of the participants rather than upon a sophisticated choice of criterion. The criterion actually chosen can be criticized from many points of view. For example, while enemy losses and our losses would clearly both be important elements in the ideal objectives function, there is no reason (and none is suggested by our authors) why one should be divided by the other. *Prima facie*, it would appear that the absolute magnitude of either loss is too important to be ignored. I will have something to say later about the shortcomings of ratios as criteria in operations problems. What is far more important in this case is the complete neglect of another dimension of the objectives function which appears to an outsider to be as important as those considered--viz.,

^{1/} Italics the authors'.

^{2/} Italics mine.

the reduced operating efficiency of ships in large convoys, and hence the inverse relation between the size of convoy and the capacity of any given number of merchant ships to transport men and materiel across the Atlantic. It is not true that the case for large convoys is a persuasive one "from any point of view." Collecting large convoys takes time. The arrival of large convoys swamps port facilities, which means longer turnaround times. Because the speed of a convoy cannot exceed that of the slowest ship, there will be an inverse average relation between its size and speed. It might well be worth a few additional sinkings to insure the delivery in time of the forces required for the Normandy invasion. The complete omission of this objectives dimension is curious because it is so admirably adapted to analysis by quantitative methods. Presumably the explanation is that a quantitative analysis had already been made (perhaps by others) of the effect of convoy size on the carrying capacity of the merchant fleet, and the commander was therefore able to weigh (if only in some intuitive manner) the gain and the cost of marginal increments in convoy size.

That something was wrong with their plausible criterion should have been immediately evident to the authors: it proves far, far too much. It shows that it would be desirable to increase the size of convoys without limit--until the whole merchant fleet and all the destroyers are assembled in a single convoy. The authors, it is true, warn that the equation cannot be expected to be valid for "very small" and "very large" values--but this is a conventional warning against extrapolating functions far beyond the range of the data from which they are derived. The important point is that, long before the whole Atlantic fleet becomes a single convoy, the significant reductions in losses will have been achieved and the reduction in the efficiency of utilization of shipping will have become unacceptable.

It will always be necessary to use judgment and good sense in applying the results of operations research, but we must try to find criteria which place a less overwhelming burden on these qualities.

There is, parenthetically, one other moral I wish to draw from this example before leaving it. The authors conclude, we have seen, that the results of their recommended action were even more successful than their equations had predicted, because the U-Boat fleet was withdrawn and sent elsewhere on other missions.

This is really a case of taking one's sub-optimization criterion too seriously. By that criterion^{1/} results were better than predicted, but if we look at a higher level criterion--say, effect on probability of winning the war--it is certain that Allied operations elsewhere were adversely affected by the diversion of the U-Boat fleet. Moreover, presuming that the Germans made a rational decision, their U-Boats, or the resources going into them, made a more significant contribution to German prospects of victory in the war elsewhere--after enlargement of the convoys--than they could have made by continuing to operate in the North Atlantic. In terms of the higher criterion, the effect on the probability of winning the war of taking the recommended action was less than one would infer from the calculation of results in the North Atlantic, which was based on the assumption that enemy U-Boat tactics and deployment would remain unchanged. For when we change our operations, different tactics and deployment become optimal for the enemy. By adopting them he can, in general, reduce his loss, as he did on this occasion.

The point is an important one--in many cases more important than it probably was in this, for the North Atlantic, while not the only shipping area, was by all odds the most vital one within easy reach of German U-Boats. A type of case in which it can be of dominant importance is where we are defending some operation

^{1/} Was it in this case the exchange ratio or the absolute number of merchant ships lost on the North Atlantic?

or function against enemy attack, and the operation is vulnerable to the destruction of several alternative target systems. Suppose, for example, we are defending a railway net against air attack. We decide that the net is most vulnerable to attack on its bridges, and carry out our operations analysis to determine tactics and deployment which minimize damage to bridges. Of course, if we were so foolish, we would almost certainly discover that results were "even better" than predicted in terms of bridges damaged. For the enemy would direct his attack not at the defended bridges, but at the undefended tunnels or rolling stock or open lines.

The moral in this type of case is that the low level sub-optimization criterion is not good enough. Effects must be assessed at least at the next higher level, in terms of the operation or function which we are defending. And some capacity for rational adjustment to our tactics must be attributed to the enemy--using, if not the analytic methods of game theory, some of its concepts and spirit.

Sub-optimization in Economic Theory

The only discipline I know which has made any attempt to explore the characteristics of operations criteria, and the intimately related question of the relation between lower and higher level sub-optimization, is economic theory. It has done so using a very different, and in some respects unfortunate, terminology, and has of course largely confined its attention to a particular context --that of the economy. I believe, however, that some of its conclusions and insights have wide applicability to operations research outside this context, and indeed constitute the beginnings, although very modest beginnings, of a scientific analysis of the problem of selecting operations criteria. The mathematical content of this analysis is simple,^{1/} but it embodies, I think, much practical common sense. In these respects it matches other aspects of operations research.

^{1/} See the Mathematical Note at the end of the paper.

The relevant portions of economic theory are directed to the analysis of the consequences for production in the economy of maximizing or optimizing behavior on the part of individuals and firms in the economy. It is therefore an analysis of the relations between sub-optimizing at two levels--a lower and a higher--and specifically of the consequences in terms of a higher level criterion of actions dictated by alternative lower level criteria.

An important conclusion of this branch of economics is that, on certain assumptions which we need not examine here, profit maximizing behavior on the part of individuals and firms results, for any given level of employment of resources, in an "efficient" organization of production in the economy--in the precisely defined sense that it is impossible to produce more of any single good or service without producing less of some other.^{1/}

What I want to do in the balance of this paper is to suggest a number of ways in which this economic theory has been useful to me, and I think can be generally useful, in dealing with certain problems in operations research.

1. The criterion for "good" criteria in operations research is always consistency with a "good" criterion at a higher level. All my other propositions are subordinate to this one. This appeal to higher levels involves no circular chasing of mirages, but acceptance, at some level of optimization, of an authoritative or self-evident statement of objectives. The test of the profit maximization criterion in the firm's sub-optimizing is the effect on production in the economy. In our example, the test of the exchange rate between U-Boats and M/V's as a sub-optimization criterion in the convoy problem was consistency with the higher level criterion of probability of victory: and there proved to be, in the general case, no necessary connection.

^{1/} Leisure is counted, rightly, as an economic good.

2. Where, for practical reasons of convenience, a sub-optimization criterion must be used which is known to be inconsistent with a higher level criterion, allowance must be made for gains or losses imposed on other operations related to the higher level criterion.

The profit maximization criterion results in maximum production in the economy only if, to use economists' jargon, "social" product equals "private" product, and "social" cost equals "private" cost. If, in draining my field, I necessarily drain my neighbor's too, social product exceeds private. If the smog generated by my refining operations imposes costs on the city of Los Angeles or any of its citizens, social costs exceed private. Divergences of this kind in either direction lead to uneconomic use of resources--from the point of view of the economy as a whole.

Examples in all types of operations research are a dime a dozen. It would apparently have been easy to win the Battle of the Atlantic against the U-Boats in a way which would have imperilled Overlord. Certain methods of carrying out Air Force missions impose special burdens on the Army or Navy or other Air operations: or alternatively, relieve them of burdens previously assumed. In industrial operations related to a single operation, process, product or plant of a company, it is almost inevitable that solutions will affect, favorably or adversely, other operations of the company.

Occasionally it is possible to take divergences between private and social (or more appropriately in operations research, low and higher) criteria explicitly into account in the analysis. An example is the problem discussed in a recent Fortune article on Operations Research^{1/}. In optimizing production and inventory policy for a single product we will almost certainly affect the costs of other products produced by a firm; and possibly in ways which can readily be estimated, and once estimated, included in the criterion, since all the costs have a common

^{1/} Herbert Solow, "Operations Research," Fortune, Vol. XLIII, No. 4 (April 1951), pp. 105-107.

monetary dimension. More frequently, as in the convoy example, there will be divergences between local or low level and high level effects about which we can know only the direction, or at most the magnitude along some dimension not commensurable with those in our local criterion. Good decisions cannot ignore such divergences, but must take them into account intuitively.

3. Ratios are particularly treacherous as operations criteria whether they are Morse-Kimball exchange rates (enemy losses divided by ours) or, more generally, ratios of objectives achieved to (in some sense) costs incurred.

In the economy production is maximized (in the sense which I previously defined) if firms maximize their profits in an absolute sense; i.e., gross receipts minus costs. It might appear as plausible, or more so, that the economy would be most efficiently organized if each firm minimized cost per unit of output, or maximized profits per unit of output. But it can be demonstrated that both of these sub-optimization criteria will result in inefficient utilization of resources by the economy.

I think there are two interacting reasons why ratios tend to make treacherous criteria both in economics and in operations problems:

- (a) They ignore the absolute magnitudes of both numerator and denominator.
- (b) Solutions with ratio criteria tend, in many operations problems, to rush to corners--for example, in the convoy case, to the corner in which all vessels are assembled in a single convoy. This may be because the ratios are really maximized or minimized in corners, or because, for simplicity, linear functional forms are assumed. But it is precisely in rushing to a corner that the absolute magnitudes cannot be ignored.

If the scale of the operation is given either in terms of resources or budget available, or in terms of a precise objective to be accomplished, then it does no harm to maximize the ratio of objectives to costs. This is mathe- matically equivalent to maximizing the objectives function subject to resource constraints, or to minimizing the costs of achieving a given objective. But if the scale of the operation is not given, i.e., if both costs and objectives are permitted to vary freely, the calculation of an optimum ratio between them will suggest a scale for the operation which bears only an accidental relation to any higher level criterion.

4. The most common and fatal mistake in selecting sub-optimization criteria is to concentrate on a single input--to maximize some objectives function for a given quantity of the input, or minimize requirements for the input to achieve some given objective. The fallacy, in the jargon of economic theory, consists in treating all other inputs as if they were free goods.^{1/}

Let me take my example here directly from economics. There has been a great deal of excitement in recent years regarding productivity per head or per man-hour in different industries in different countries. Maximum output per head (or man-hour) seems so obviously desirable that most non-economists would never question it as a suitable criterion for organizing production. In fact, as a sub-optimizing criterion for any firm or industry, it is quite wrong by our only test--the next higher level criterion. If firms or industries use it for sub-optimizing the resulting organization of production in the economy will be inefficient.

The reason is obvious on reflection. There are other scarce, valuable resources needed in production in addition to man-hours, and if they are not

^{1/} In a different jargon, the fallacy is to employ a partial or incomplete criterion without recognizing its incompleteness.

used "economically," overall production will suffer. These other resources include capital, land, management, and labor with special skills or training. A method of production chosen to minimize man-hours is likely to be inordinately wasteful of these other valuable resources. A bombing system designed to minimize requirements for bombs will be unnecessarily and inefficiently (from the point of view of the higher level criterion) wasteful of aircraft and crews. Conversely, if it is designed to save aircraft and crews, it will waste bombs. "Hemibel thinking" ^{1/} does not come to our rescue here: systems which minimize input A frequently turn out to be utterly different from systems which minimize input B (as in the case illustrated in Fig. 1 below).

5. The "Golden Rule" for allocating scarce resources is to make each resource equally scarce in all uses. This is a theorem of "welfare economics," with precisely and operationally defined terms, which enables operations researchers to give practical advice on allocation problems even where the values of the alternative uses are incommensurable in the higher level objectives function.

An objective can usually be achieved by various combinations of resources. Motor cars can be made with much capital and little labor, as in Detroit, or more labor and less capital, as in Coventry. Agricultural crops can be grown using intensive or extensive methods--requiring very different proportions of land, labor, and machinery. An air defense of any specified effectiveness may be achieved with very different mixes of interceptors and anti-aircraft artillery. Some substitutions of resources are direct and obvious; many others are indirect and subtle. The operations researcher typically discovers that the possibilities of substitution in carrying out any operation are much greater than are at first apparent.

To illustrate the theorem geometrically let us first assume a case where there are only two scarce and valuable resources, A and B, which are combined to

^{1/} Morse and Kimball, op. cit., p.33.

produce a single product, P (See Fig. 1).^{1/} A might be land, B, labor and P, wheat; alternatively A might be bombers, B, bombs and P, targets destroyed. If we measure A along one axis and B along the other, we can represent production possibilities by a series of curves which will usually be convex to the origin.^{2/} Any given curve, or production "isoquant," shows the minimum combinations of resources A and B which are needed to produce a given quantity of P. The slope of the curve at any point measures the amount of B one must substitute per unit of A to maintain the output of P, and is called the "marginal rate of substitution" of B for A. Its numerical value in the normal convex case diminishes as A is substituted for B, i.e., as we move along the isoquant from left to right.

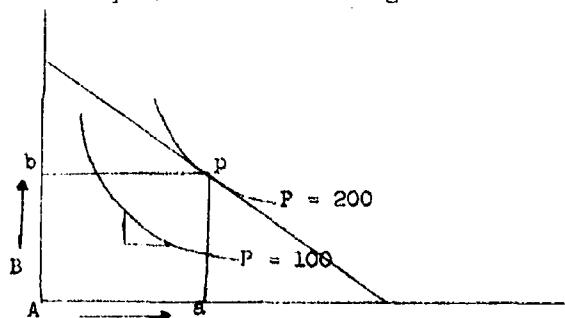


Fig. 1

For some simple types of sub-optimizations such an array of possibilities provides sufficient information for a solution. Thus, if the quantities of A and B have been fixed (e.g., for a field commander) by higher authority at a and b, the optimal operation is the one represented by point p, where all of both resources are optimally utilized to produce a P of 200. Alternatively, if what is given is the scale of the operation (e.g., $P = 200$) and the relative costs of the two resources (represented by the slope of the diagonal through p), the optimal (i.e., "least cost") operation is p, at which the relative cost diagonal is tangent to

^{1/} The P's and Q's of this and subsequent examples need not be quantities; they may simply be indices of any objectives, the degree of achieving which can be ordered.

^{2/} Convexity means simply that it becomes progressively harder to substitute A for B as the ratio of A to B increases.

the production isoquant; or, in other words, where the relative costs of the resources are equal to the marginal rate of substitution between them.

But suppose the problem is more complex. Suppose there are several objectives--either industrial products or military missions--requiring the same scarce resources. We must distinguish two cases--first, where there is a common measure for the various objectives; second, where no common measure can be found.^{1/}

The first case, where there is a common measure, turns out to be trivial, for it reduces to the simple case which we have already considered. This is the case typically found in industry, where the products of a firm can be measured in a common monetary unit. The familiar conditions for an optimum usually define a unique solution.

It is the second case, where there is no common measure, that the theorem which I have called the Golden Rule for allocating becomes a helpful device for operations researchers who might otherwise be frustrated.

Equal scarcity of each resource in each use means equal marginal rates of substitution among resources in each use.^{2/} To illustrate the significance of this geometrically let us assume only two resources, A and B, and two objectives, P and Q. A could be Robinson Crusoe's time, B the arable land on his island, and P and Q beans and squash, respectively. Alternatively, A and B could be merchant ships and destroyers, and P and Q tons of materiel moved across the Atlantic and Pacific.

First consider the production possibilities for P, measuring A and B as before, from the origin, O (Fig. 2a).

^{1/} I mean: none can be found which, as a practical matter, can be used as a measure. Knowing that a common measure exists in principle (e.g., contribution to the probability of victory) is not enough.

^{2/} I.e., among resources which it is efficient to employ in any use. It will frequently be inefficient to employ some resources in some uses (corner solutions). See the Mathematical Note.

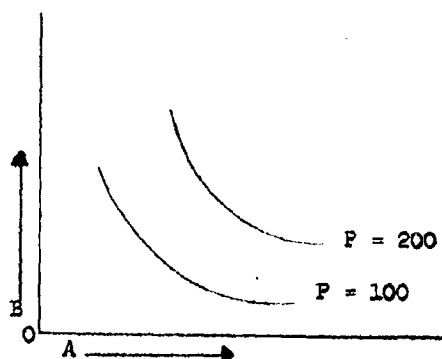


Fig. 2a

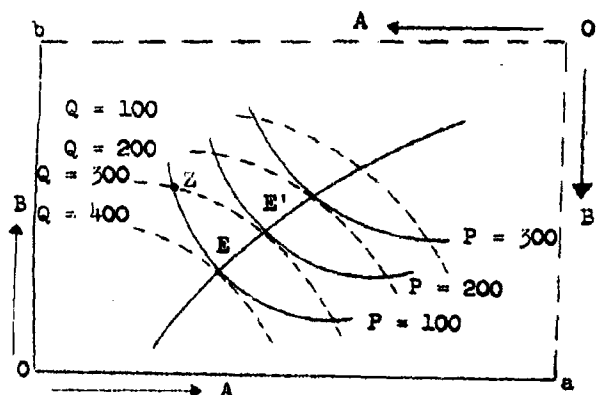


Fig. 2b

Now let Oa and Ob represent the total quantities of A and B available for both uses P and Q , and complete the rectangle (Fig. 2b).

From the origin O' , measuring B south and A west, represent the production possibilities of Q . The isoquants will (normally) be convex to the origin O' . In the usual case therefore each P isoquant will be tangent to one and only one Q isoquant.

Each point in the rectangle represents a particular allocation of the resources A and B between the objectives P and Q . The isoquants passing through that point "score" the allocation in two dimensions, P and Q . Because we have no common measure for P and Q we cannot say what is the "most preferred" of the points or the optimal allocation. But our Golden Rule permits us to say a great deal that is practically useful about bad allocations and better allocations.

Consider point Z which is, let us assume, the actual allocation or the currently planned or programmed allocation. Is it a good one? Well, it will produce 100 P and 300 Q . It is perfectly apparent that we can recommend something better. Moving along the $P = 100$ isoquant we find a point of tangency E where $Q = 400$. We do not need a common measure for P and Q to know that $(P = 100, Q = 400)$ is better than $(P = 100, Q = 300)$.^{1/} We can get to this superior point simply by transferring some A from Q to P , and some B from P to Q .

^{1/} Of course, we have to know something about the character of the higher level criterion to be sure of even this, but we do not need to know much.

Point Z is an "inefficient" point, when the marginal rates of substitution of the resources (represented by the slopes of the isoquants) are different in different uses. It is therefore possible by reallocating to get more of at least one product without having less of any other. Point E is an "efficient" point, where marginal rates of substitution are the same in all uses. It is very easy to get allocations represented by very inefficient points, especially where command responsibility is divided, and there is no system of prices to indicate relative scarcities.

Of course point E' is also an efficient point and demonstrably better than Z. So are all points of tangency between E and E', where we get more of both objectives than at Z. So we cannot recommend the optimum.

The philosophy of operations research, however, and my own, is that it is far more important to be able to demonstrate that some courses of action, E, are better than proposed courses of action, Z, than to spend one's life seeking the optimum optimorum.

Conclusion

I have not tried to give an exhaustive account of the relevancy of economics to operations research, but to suggest that it does contain the beginnings of a scientific and practically rewarding approach to the problems of choosing criteria. That approach involves the analysis of relations between sub-optimizations at lower and higher levels.

The analysis of these relations can be helpful even where the full optimization at the higher level cannot be calculated. Morse and Kimball are skeptical about the ability of operations researchers to contribute importantly to the high level problems directly, and I share their skepticism in some degree. But operations researchers must understand the general characteristics of the higher level optimization if they are to exercise good judgment in the selection of criteria at the lower levels--which means, if the sub-optimizations are to contribute even indirectly to the high level objectives.

Another example from economics may help to clarify this point. The U.S. industrial economy is very much more productive than that of the U.K.--by any crude measure you wish to use. This is so not because the U.S. government or the National Association of Manufacturers has worked out a better high level optimization for the U.S. economy than the British government or the Federation of British Industries has worked out for the British economy. Nobody attempts such a high level optimization here or in the U.K., although the Russians have a system which requires them to do so. The explanation of the difference, therefore, to the extent that it is not accounted for by our more abundant natural resources, must lie in the different way in which sub-optimizing--by individuals and firms--is done in the two countries. Specifically, the explanation must be one or both of the following:

1. The British are not as good at sub-optimizing as we are; or
2. They use inferior criteria in sub-optimizing; i.e., criteria less consistent with higher level (economy-wide) optimizing.

There is perhaps some positive evidence on both counts. I want to consider only the second. Most individuals and firms, in making economic decisions, compromise between two objectives--profits on the one hand, and security and a quiet life on the other. For reasons better left to sociologists and social psychologists to explain, the British working man and the British business man appear to give much greater weight to security than their American counterparts. Prima facie, this is neither wicked nor foolish. "If they want security and are willing to pay for it, let them have it."

But we have a theory which shows that, in a sense, it is foolish. Seeking maximum profits at this level does (under certain conditions) lead to a higher level optimization--an efficient organization of production in the economy. Seeking security at this level leads to an inefficient organization of production

in the economy--and does not even, in general, lead to a secure economy. Analogously, maximizing "exchange rates" in military operations does not, in general, promote our chances of victory.

This is the principle. This is the test of good sub-optimizing. The operations researcher will do most of his effective work on low level problems. But he will do better work if he studies and bears in mind the characteristics of the optimization at the appropriate higher level, and the relation to it of his sub-optimizing criteria.

Mathematical Note

A. W. Marshall

Of the several points discussed in the main text of the paper, perhaps only one requires any mathematical restatement or explication. This one is the so-called "Golden Rule" for the allocation of scarce resources. Here we present, for a simple case, the proof of the theorem on which it is based.^{1/}

Let us consider the case where there are several objectives, or outputs, and there is no available common measure of the value of the various outputs. Denote these outputs as X_i ; ($i = 1, \dots, k$) and the production function relating the output to the inputs as

$$(1) \quad X_i (y_{1,i}, y_{2,i}, \dots, y_{n,i}); (i = 1, \dots, k)$$

where $y_{1,i}$ is the amount of the first input or resource used in producing the i th output.^{2/} All X 's and y 's are obviously required to be non-negative. It is assumed that the total quantities of each of the resources is fixed, thus^{4/}

$$(2) \quad \sum_{i=1}^k y_{j,i} = k_j; (j = 1, \dots, n.)$$

^{1/} For a discussion of a very general case readers are referred to K. J. Arrow's paper, "An Extension of the Basic Theorem of Classical Welfare Economics," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press: Berkeley, pp. 507-532.

^{2/} In all that follows we assume that the X_i 's are differentiable functions, concave in the $y_{j,i}$'s ($j = 1, \dots, n$).

^{3/} The formulation used here is largely patterned after P. A. Samuelson's treatment of the problem in his Foundations of Economic Analysis, Harvard University Press: Cambridge, pp. 229-233.

^{4/} When the resource constraints are expressed as inequalities (\leq) rather than equalities in (2) the problem becomes one of linear or nonlinear programming depending upon the nature of the functions $X_i (y_{1,i}, \dots, y_{n,i})$. See T. C. Koopmans (ed.), Activity Analysis of Production and Allocation, Cowles Commission Monograph 13, Wiley; New York, 1950, and H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press: Berkeley, pp. 431-492.

The problem is to allocate these fixed resources so that, in some sense, the production of outputs is optimal. Since there is assumed to be no way of comparing or valuing the various outputs one against another, only a rather weak sort of optimality can be achieved. Optimal production from given resources will therefore be defined as any organization of production; i.e., allocation of the given resources to the production of the various outputs, such that no other organization of production will yield a greater quantity of one output without a concomitant reduction in the output of some other output or outputs. In other words we are concerned with a partial ordering of the vectors X of possible outputs, given the fixed resources available. This partial ordering will separate the set of possible vectors into two distinct subsets: an efficient set and an inefficient set. The set of inefficient vectors is defined by the condition, X is a member of the set of inefficient vectors of production if there exists a producible vector $X^* \neq X$ such that $X \leq X^*$, where the relations \leq , $<$, etc., hold as though written for each component individually.^{1/} The set of efficient vectors may thus be obtained in the following way: Arbitrarily assign outputs O_1 of $(k - 1)$ of the k outputs and maximize the remaining output subject to the $(k - 1)$ production and n resource constraints. The set of all vectors obtained in this way for all possible arbitrary assignments of the $(k - 1)$ output constraints is the set of efficient vectors. For simplicity it will be assumed that all solutions to this maximization problem are interior solutions in the space of the $y_{i,j}$'s (Arrow treats the more

^{1/} If S_1 and S_2 denote the sets of inefficient and efficient vectors of production, respectively, and $S_1 + S_2 = S$, $= S$, the whole space, then inefficient vectors are defined by

$$X \in S_1 \equiv \exists X^* \neq X \ni X^* \geq X$$

or alternatively we may define efficient vectors by

$$X \in S_2 \equiv \nexists X^* \neq X \ni X^* \geq X.$$

general case).^{1/} The problem is most easily dealt with by use of the method of Lagrange multipliers which leads to the formulation:

Maximize

$$\bar{\Phi} = X_k(y_{1,k}, \dots, y_{n,k}) + \sum_{i=1}^{k-1} \lambda_i \left[X_i(y_{1,i}, \dots, y_{n,i}) - O_i \right] \\ + \sum_{j=1}^n \left\{ \pi_j \left[\sum_{i=1}^k y_{j,i} - k_j \right] \right\}$$

with side conditions

$$X_i(y_{1,i}, \dots, y_{n,i}) - O_i = 0; (i = 1, \dots, k-1)$$

$$\sum_{i=1}^k y_{j,i} - k_j = 0$$

After taking the partial derivatives of $\bar{\Phi}$ with respect to the $y_{i,j}$'s and eliminating the Lagrange multipliers, the first order maximum conditions take the form^{2/}

$$\frac{\frac{\partial X_1}{\partial y_{1,i}}}{\frac{\partial X_k}{\partial y_{1,k}}} = \dots = \frac{\frac{\partial X_1}{\partial y_{n,i}}}{\frac{\partial X_k}{\partial y_{n,k}}} ; (i = 1, \dots, k-1)$$

^{1/} When a corner solution does result; i.e., some $y_{i,j}$'s are equal to zero, the marginal productivity of inputs i in use j must be less than or equal to its marginal productivity in all other production processes in which it is actually being used.

^{2/} In order to verify that this is a true maximum the second order conditions will of course always have to be investigated.

Since the choice of the maximization of the k th output was arbitrary it is clear that both i and k can range over the indices $1, \dots, k$ and the first conditions for the maximum remain the same. An equivalent form of these conditions is

$$\frac{\frac{\partial x_1}{\partial y_{i,1}}}{\frac{\partial x_1}{\partial y_{j,1}}} = \dots = \frac{\frac{\partial x_k}{\partial y_{i,k}}}{\frac{\partial x_k}{\partial y_{j,k}}} ; (i, j = 1, \dots, n)$$

In words this takes the form: Production factors are correctly (efficiently) allocated if the ratio of the marginal productivity of a given factor in one use to the marginal productivity of the same factor in a second use is the same as the ratio of the marginal productivity of any other factor in the first use to its marginal productivity in the second use.

The statement of the "golden rule" in the text is in terms of marginal rates of substitution of resources to produce some fixed output rather than in terms of marginal productivities. Since marginal rates of substitution are related to marginal productivities by the relation

$$\frac{dy_{i,h}}{dy_{j,h}} = - \frac{\frac{\partial x_h}{\partial y_{j,h}}}{\frac{\partial x_h}{\partial y_{i,h}}} ; (i, j = 1, \dots, n \text{ and } h = 1, \dots, k)$$

it is clear that when the marginal productivities of each input are the same in all outputs or uses the marginal rates of substitution between inputs in the production of each of the outputs will be equal, if the allocation is an efficient one.